



A COMPARATIVE ANALYSIS OF UNCERTAIN QUERY PROCESSING USING FUZZY SETS AND VAGUE SETS

Jaydev Mishra

Department of Computer Science and Engineering, College of Engineering & Management, Kolaghat, West Bengal, India.

ABSTRACT

The most important aspect in the utilization of a database system is its ability of processing information and queries correctly. The objective of the present paper is to analyze and compare the performance of fuzzy and vague database models with respect to processing of uncertain queries. An algorithm has been designed for that purpose and has been successfully applied to queries related to real life examples. The study reveals that vague sets produce more accurate decisions in comparison to fuzzy sets and thus a DBMS that uses vague theoretic concept may become a more powerful software product than those currently available.

KEYWORDS: Fuzzy set, Vague set, Similarity measure, SQL

1. INTRODUCTION:

Information in the real world is often incomplete or imprecise in nature. The traditional query in a relational database system is unable to deal with such imprecise data. It is widely recognized that fuzzy set theory, proposed by Zadeh [1] in 1965, has long been introduced to handle inexact and imprecise data. Several authors [2-6] have also provided a theoretical foundation to query language for a fuzzy data model. In particular, the well known SQL language has been extended by Bosc et al. [4] and Nakajima [5] in the framework of fuzzy set theory to develop a fuzzy SQL language, called SQLF. In ref. [6] authors have recently stated an approach where users can retrieve data without any prior knowledge of the database schema or formal query language. The efficiency as well as benefits of using fuzzy query over crisp query in a classical database has been illustrated in [7].

The theory of vague sets was introduced by Gau and Buehrer [8] in 1993 as a generalization of fuzzy sets. A vague set uses interval-based membership values and is believed to have more powerful ability to process imprecise information than fuzzy sets. The notion of vague sets has been incorporated into relations in [9] and a vague SQL, termed as VSQL, has been developed. Zhao and Ma [10] have also proposed a vague database model and have investigated vague querying strategies with SQL. In [11], Dutta et al. have adopted vague set as a tool to generate a new method of intelligent search, called "vague search" that would answer any imprecise query set by the database user. All the design aspects of a classical relational database model have been thoroughly extended by Mishra [12] in the light of vague set theory which has led to the development of a new vague relational database model. Mishra et al. [13] have also designed architecture to process uncertain queries and have utilized the same to make a theoretical comparison between fuzzy sets and vague sets in processing uncertain queries.

The objective of the present paper is also to make a comparative analysis of uncertain query processing using fuzzy sets and vague sets. It has been observed that the algorithm presented in [13] for finding membership values has certain drawbacks. In this work, a new algorithm has been proposed that defines an attribute independent membership function of a fuzzy/vague set for calculating membership value of different fuzzy or vague attributes and the algorithm is free from any such anomalies. Also, it may be worthwhile to note that membership functions used previously in literature [14,15] are found to be dependent on the particular attribute whereas the present algorithm generates membership values which is independent of the type of the attribute.

The organization of the paper is as follows: Basic definitions related to fuzzy set and vague set are presented in Section 2. The definition of similarity measure between two vague data also appears in the same section. In Section 3, an algorithm has been designed to get the appropriate membership value and represent domain value of fuzzy or vague attributes into fuzzy form or vague form. Complexity analysis of the proposed algorithm is presented in same section. With the help of real life examples, it is observed in Section 4 that a vague set is more appropriate than fuzzy set in processing imprecise queries. The final conclusion of the paper is reported in Section 5.

2. BASIC DEFINITIONS:

Some basic definitions related to fuzzy and vague sets as well as a similarity measure between two vague sets used in the paper is presented in this section. Let U be the universe of discourse where an element of U is denoted by u .

2.1 Fuzzy Set

Definition 2.1.1 A fuzzy set F in a universe of discourse U is characterized by a membership function $\mu_F : U \rightarrow [0,1]$ and is defined as a set of ordered pairs

$$F = \{(u, \mu_F(u)) : u \in U\}, \text{ where } \mu_F(u) \text{ for each } u \in U \text{ denotes the grade of membership of } u \text{ in the fuzzy set } F.$$

It may be noted that a classical subset A of U can be viewed as a fuzzy subset with membership function μ_A taking binary values, i.e.

$$\mu_A = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{if } u \notin A \end{cases}$$

2.2 Vague Set:

Definition 2.2.1 A vague set V in the universe of discourse U is characterized by two membership functions given by:

(i) a truth membership function $t_V : U \rightarrow [0,1]$, and

(ii) a false membership function $f_V : U \rightarrow [0, 1]$,

where $t_V(u)$ is a lower bound of the grade of membership of u derived from the ‘evidence for u ’, and $f_V(u)$ is a lower bound on the negation of u derived from the ‘evidence against u ’, and $t_V(u) + f_V(u) \leq 1$.

Thus, the grade of membership $\mu_V(u)$ of u in the vague set V is bounded by a subinterval $[t_V(u), 1 - f_V(u)]$ of $[0, 1]$ i.e., $t_V(u) \leq \mu_V(u) \leq 1 - f_V(u)$.

Then, the vague set V is written as $V = \{ \langle u, [t_V(u), 1 - f_V(u)] \rangle : u \in U \}$ and the interval $[t_V(u), 1 - f_V(u)]$ is said to be the vague value to the object u . The precision of knowledge about u is clearly characterized by the difference $(1 - f_V(u) - t_V(u))$. If this is small, the knowledge about u is relatively precise. However, if it is large, we know correspondingly little. If $t_V(u)$ is equal to $(1 - f_V(u))$, the knowledge about u is precise, and vague set theory reverts back to fuzzy set theory. If $t_V(u)$ and $(1 - f_V(u))$ are both equal to 1 or 0, depending on whether u does or does not belong to V , the knowledge about u is very exact and the theory goes back to that of ordinary sets. Thus, any crisp or fuzzy set may be considered as a special case of vague set.

2.3 Similarity Measure:

There have been several studies in literature regarding the degree of similarity between vague sets [16-19]. It was shown by Lu et al. in ref. [19] that the similarity measures defined in [16-18] did not work well in some cases. They have proposed a new similarity measure between vague sets which has been proved to be reasonable in more general cases. The same has been utilized in the present study and is defined as follows:

Definition 2.3.1 *Similarity Measure between two vague values*

Let X and Y be any two vague values such that $X = [t_x, 1 - f_x]$ and $Y = [t_y, 1 - f_y]$, where $0 \leq t_x \leq 1 - f_x \leq 1$, and $0 \leq t_y \leq 1 - f_y \leq 1$.

Let $SE(X, Y)$ denote the similarity measure between X and Y . Then,

$$SE(X, Y) = \sqrt{\left(1 - \frac{|(t_x - t_y) - (f_x - f_y)|}{2}\right) \left(1 - |(t_x - t_y) + (f_x - f_y)|\right)}$$

Definition 2.3.2 *Similarity Measure between two vague sets*

Let $U = \{u_1, u_2, u_3, \dots, u_n\}$ be the universe of discourse. Let A and B be two vague sets on U , such that

$$A = \{ \langle u_i, [t_A(u_i), 1 - f_A(u_i)] \rangle, \forall u_i \in U \},$$

Where $t_A(u_i) \leq \mu_A(u_i) \leq 1 - f_A(u_i)$ and $1 \leq i \leq n$.

$$B = \{ \langle u_i, [t_B(u_i), 1 - f_B(u_i)] \rangle, \forall u_i \in U \},$$

Where $t_B(u_i) \leq \mu_B(u_i) \leq 1 - f_B(u_i)$ and $1 \leq i \leq n$.

Now, the similarity measure between A and B , denoted by $SE(A, B)$ is defined as:

$$SE(A, B) = \frac{1}{n} \sum_{i=1}^n SE([t_A(u_i), 1 - f_A(u_i)], [t_B(u_i), 1 - f_B(u_i)])$$

$$= \frac{1}{n} \sum_{i=1}^n \sqrt{\left(1 - \frac{|(t_A(u_i) - t_B(u_i)) - (f_A(u_i) - f_B(u_i))|}{2}\right) \left(1 - |(t_A(u_i) - t_B(u_i)) + (f_A(u_i) - f_B(u_i))|\right)}$$

3. ALGORITHM FOR FINDING MEMBERSHIP VALUE:

For finding membership value, different membership functions have been defined by several authors in literature for different fuzzy attributes. For example, in the relation **EMP (Name, Age, Experience, Salary)**, Raju et. al. [14] and Ma et. al. [15] have defined membership functions for the fuzzy set ‘close to u ’ for the two different fuzzy attributes **Salary** and **Experience** as follows:

According to Raju et. al. [14], membership function for the fuzzy set ‘close to u ’ for fuzzy attribute **Salary** is defined as:

$$\mu_{close\ to\ u}(x) = 1.0 / (1 + |x - u| / 20000)$$

whereas the membership function for fuzzy set ‘close to u ’ for fuzzy attribute **Experience** is defined as:

$$\mu_{close\ to\ u}(x) = 1.0 / (1 + |x - u|/4)$$

However, the membership functions defined by Ma et. al. [15] for the fuzzy attributes *Salary* and *Experience* are respectively given by

$$\mu_{close\ to\ u}(x) = 1.0 / (1 + (|x - u|/20000)^2)$$

and

$$\mu_{close\ to\ u}(x) = 1.0 / (1 + (|x - u|/4)^2).$$

From the above observations, it may be clearly noted that the membership functions defined by Raju et al. as well as Ma et al. are dependent on the particular attribute under consideration.

In the present work, an effort has been taken to develop an algorithm which will generate an attribute independent membership function of a fuzzy set or vague set for calculating membership values of different fuzzy/vague attributes. The proposed algorithm yields similar membership values as compared with other authors.

Further, it may be noted that Mishra et al. have also presented a similar algorithm in ref. [13] for calculation of membership values. However, the algorithm proposed in [13] has the following drawbacks:

- (i) In certain cases, the algorithm in ref. [13] yields negative membership value.
- (ii) If new records are added in the given data, the membership values of the existing data also change which is not desirable.

The algorithm proposed in the present paper is shown below which is free from the anomalies mentioned above.

Algorithm_1

Input: Fuzzy / Vague attributes and Fuzzy / Vague data given in the uncertain query.

Output: A membership value in the interval [0, 1].

Method:

find the fuzzy or vague attributes from the uncertain query.

for each fuzzy or vague attribute **do**

begin

fdata \leftarrow data value for the fuzzy or vague attribute of the query

range = maxDomValue – minDomValue

for each tuple of the relation **do**

begin

tvalue \leftarrow corresponding tuple value from the fuzzy or vague attribute domain

if ($|fdata - tuple_value| \leq range$)

membershipValue = $1 - (|fdata - tvalue| / range)$

else membershipValue = 0

end for loop for tuple

end for loop for fuzzy or vague attribute.

3.1 Complexity analysis of algorithm_1 for processing uncertain query:

While processing any uncertain query from a relation r having m attributes and n tuples, our proposed algorithm finds the membership values for each tuple with respect to fuzzy/vague attributes present in the uncertain query. In the worst case any uncertain query may contain all fuzzy/vague attributes of the relation r . Here we compute the worst case time complexity of the Algorithm 1.

Lemma 1: Worst case time complexity for processing any uncertain query from a relation r having m attributes and n tuples is $O(mn)$.

Proof: In the worst case, any uncertain query contains all m attributes as fuzzy/vague data. In such a situation the outer loop will run m times. Again for each iteration of the outer loop there are some statesman which runs in constant time i.e., $O(1)$ and it also contains an inner loop which runs for n tuples with complexity $O(n)$. So, for each iteration, outer loop takes $O(n) + O(1)$ i.e., $O(n)$. Hence for m iterations time complexity of our proposed algorithm is $O(mn)$.

4. REAL LIFE APPLICATION: VAGUE SETS ARE BETTER THAN FUZZY SETS

Consider the following Employee (*EMP*) relation and the uncertain queries:

Table 1: EMP relation

Name	Age (yrs)	Exp (yrs)	Sal (Rs)
Mr. Smith	25	2	30000
Mr. Ganguly	28	4	30500

Mr. Roy	31	6	32000
Mr. David	48	16	56000
Mr. Maity	54	20	75900
Mr. Das	46	15	55400
Mr. Ahuja	47	14	54800
Mr. Sharma	38	11	44500
Mr. Kundu	36	10	40000
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Bose	53	19	79100
Mr. Gayen	51	21	80200
Mr. Guha	52	25	80000
Mr. Ghosh	59	33	120000

Uncertain query 1: “To find the data of the Employees whose age is *about* 50”.

Uncertain query 2: “To find the details of the Employees whose age is *more or less* 50 and experience is *near* 20”.

The above two queries have been executed both by the fuzzy as well as by the vague database model and each time it is observed that the vague database model gives better output in comparison to the fuzzy database model. The details are as follows:

Uncertain query 1: “To find the data of the Employees whose age is *about* 50”.

Solution using Fuzzy Set:

Here, our fuzzy attribute is *Age* and the fuzzy data is *about* 50.

Let us now apply the **Algorithm_1** to fetch the membership value corresponding to every domain value of fuzzy attribute *Age*.

Input

Fuzzy attribute *Age*, fuzzy data *about* 50.

Method

$\text{dom}(\text{Age}) = \{25, 28, 31, 48, 54, 46, 47, 38, 36, 50, 49, 53, 51, 52, 59\}$

given $\text{fdata} = 50$

$\text{range} = 59 - 25 = 34$

Now, we use the formula specified in the **Algorithm_1** to calculate membership values for each domain value of *Age*. Thus,

for the 1st tuple : $\text{membershipValue} = 1 - (|50-25| / 34) = 0.26$

for the 2nd tuple : $\text{membershipValue} = 1 - (|50-28| / 34) = 0.35$ and so on.

The entire list of membership values for all tuple values of the fuzzy attribute *Age* for the fuzzy set ‘*about* 50’ is given below:

Table 2: Membership Values

Age (yrs)	Membership Value
25	0.26
28	0.35
31	0.44
48	0.94
54	0.88
46	0.88
47	0.91
38	0.65
36	0.59
50	1
49	0.97
53	0.91
51	0.97
52	0.94
59	0.74

The fuzzy representation of the above relation with respect to the **uncertain query 1** is now shown in **Table 3**. Here, the fuzzy representation of the *Age* attribute occurs in third column while the corresponding vague representation is depicted in the fourth column. Now the vague values are used to calculate the similarity measures (S.M.) with fuzzy data ‘*about* 50’ whose vague representation is $\langle 50, [1, 1] \rangle$. The similarity measure formula as presented in **Definition 2.3.1** has been used as follows. If we consider the two vague data $x = \langle 50, [1, 1] \rangle$ and $y = \langle 25, [.26, .26] \rangle$, then $t_x = 1$, $f_x = 0$, $t_y = .26$, $f_y = .74$, and thus

$$\begin{aligned}
 S.M.(x, y) &= \sqrt{\left(1 - \frac{|(1-.26) - (0-.74)|}{2}\right)} \left(1 - |(1-.26) + (0-.74)|\right) \\
 &= \sqrt{(1-.74)} = \sqrt{.26} = 0.51
 \end{aligned}$$

Again, for $x = \langle 50, [1, 1] \rangle$ and $y = \langle 28, [.35, .35] \rangle$, $t_x = 1$, $f_x = 0$, $t_y = .35$, $f_y = .65$. Then,

$$S.M.(x, y) = \sqrt{\left(1 - \frac{|(1-.35) - (0-.65)|}{2}\right)} \left(1 - |(1-.35) + (0-.65)|\right)$$

$$= \sqrt{(1-.65)} = \sqrt{.35} = 0.59$$

and so on. The complete table is given below

Table 3: Similarity Measures of tuples of EMP relation with respect to uncertain query 1 using fuzzy set

Name	Age	Fuzzy representation of Age with fuzzy data about 50	Vague representation for fuzzy attribute Age	S.M. of Age with <50,[1,1]>	Exp	Sal	S.M.(tuple)
Mr. Smith	25	<25, .26>	<25, [.26,.26]>	.51	2	30000	.51
Mr. Ganguly	28	<28, .35>	<28, [.35,.35]>	.59	4	30500	.59
Mr. Roy	31	<31, .44>	<31, [.44,.44]>	.66	6	32000	.66
Mr. David	48	<48, .94>	<48, [.94,.94]>	.97	16	56000	.97
Mr. Maity	54	<54, .88>	<54, [.88,.88]>	.94	20	75900	.94
Mr. Das	46	<46, .88>	<46, [.88,.88]>	.94	15	55400	.94
Mr. Ahuja	47	<47, .91>	<47, [.91,.91]>	.95	14	54800	.95
Mr. Sharma	38	<38, .65>	<38, [.65,.65]>	.81	11	44500	.81
Mr. Kundu	36	<36, .59>	<36, [.59,.59]>	.77	10	40000	.77
Mr. Dutta	50	<50, 1>	<50, [1,1]>	1	18	78900	1
Mr. Mondal	49	<49, .97>	<49, [.97,.97]>	.98	17	78200	.98
Mr. Bose	53	<53, .91>	<53, [.91,.91]>	.95	19	79100	.95
Mr. Gayen	51	<51, .97>	<51, [.97,.97]>	.98	21	80200	.98
Mr. Guha	52	<52, .94>	<52, [.94,.94]>	.97	25	80000	.97
Mr. Ghosh	59	<59, .74>	<59, [.74,.74]>	.86	33	120000	.86

Since the above uncertain query contains only one fuzzy attribute, hence the similarity measures of the tuples is same as the similarity measures obtained for the attribute *Age*. Now, to execute the query at a **threshold** or α -cut value, say, **0.95**, which is supplied by the decision maker, the following **SQL** statement is generated:

SELECT * FROM *EMP* WHERE *S.M.(tuple)* \geq 0.95

This retrieves the resultant tuples from the *EMP* database as shown in Table 4.

Table 4: Resultant tuples of EMP relation after processing uncertain query 1 using fuzzy set at threshold value $\alpha = 0.95$

Name	Age	Exp	Sal
Mr. David	48	16	56000
Mr. Ahuja	47	14	54800
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Bose	53	19	79100
Mr. Gayen	51	21	80200
Mr. Guha	52	25	80000

Solution using Vague Set:

Let us now study the same **query** with vague sets under consideration. Once again, the vague attribute is *Age* and vague data is *about 50*.

As we proceed to represent all domain values of the vague attribute *Age* into vague data, the truth membership values are obtained using the **Algorithm 1** while false membership values are provided randomly by the decision maker under the restriction that sum of truth and false membership values must be less or equal to 1. Similarity measures are now calculated using the same formula as used for fuzzy attributes. The complete table using vague representation is as shown below:

Table 5: Similarity Measure of tuples of EMP relation with respect to uncertain query 1 using vague set

Name	Age	Vague representation of vague attribute Age for vague set 'about 50'	S.M. of Age with <50,[1,1]>	Exp	Sal	S.M.(tuple)
Mr. Smith	25	<25, [.26,.33]>	.51	2	30000	.51
Mr. Ganguly	28	<28, [.35,.39]>	.59	4	30500	.59
Mr. Roy	31	<31, [.44,.49]>	.66	6	32000	.66
Mr. David	48	<48, [.94,.97]>	.96	16	56000	.96
Mr. Maity	54	<54, [.88,.95]>	.92	20	75900	.92
Mr. Das	46	<46, [.88,.94]>	.92	15	55400	.92
Mr. Ahuja	47	<47, [.91,.96]>	.94	14	54800	.94
Mr. Sharma	38	<38, [.65,.67]>	.80	11	44500	.80
Mr. Kundu	36	<36, [.59,.60]>	.77	10	40000	.77
Mr. Dutta	50	<50, [1,1]>	1	18	78900	1
Mr. Mondal	49	<49, [.97,.98]>	.98	17	78200	.98
Mr. Bose	53	<53, [.91,.97]>	.94	19	79100	.94
Mr. Gayen	51	<51, [.97,.99]>	.98	21	80200	.98
Mr. Guha	52	<52, [.94,.98]>	.96	25	80000	.96
Mr. Ghosh	59	<59, [.74,.78]>	.85	33	120000	.85

The following resultant tuples are now retrieved from the *EMP* database at the same α -cut value **0.95**

Table 6: Resultant tuples of *EMP* relation after processing uncertain query 1 using vague set at threshold value $\alpha = 0.95$

Name	Age	Exp	Sal
Mr. David	48	16	56000
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Gayen	51	21	80200
Mr. Guha	52	25	80000

We may interpret from the **Tables 4** and **6** that vague set theory is more effective than its fuzzy counterpart since the **SQL** statement with vague query does not fetch the tuples of Mr. Ahuja of age 47 and Mr. Bose of age 53 that have been retrieved with the fuzzy query. It is obvious that both 47 and 53 are less close to 50 compared to the values of the attribute *Age* in all the other tuples retrieved by the **SQL** statement.

Next we consider another uncertain query that contains more than one attribute of fuzzy or vague nature.

Uncertain query 2: “To find the data of the Employees whose age is *more or less* 50 and experience is *near* 20”.

Solution using Fuzzy Set:

The present query has two fuzzy attributes, namely, *Age* and *Exp*. The fuzzy representation of the *EMP* relation, as obtained by using **Algorithm_1** and **Definition 2.3.1** is presented in **Table 7**.

Table 7: Similarity Measure of tuples of *EMP* relation with respect to uncertain query 2 using fuzzy set

Name	Age	Fuzzy rep. of Age with fuzzy set more or less 50	Vague rep. of fuzzy data of fuzzy attribute Age	μ_1 S.M. of Age with $\langle 50, [1,1] \rangle$	Exp	Fuzzy rep. of Exp with fuzzy set near 20	Vague rep. of fuzzy data of fuzzy attribute Exp	μ_2 S.M. of Exp with $\langle 20, [1,1] \rangle$	Sal	μ
Mr. Smith	25	$\langle 25, .26 \rangle$	$\langle 25, [.26, .26] \rangle$	0.51	2	$\langle 2, .42 \rangle$	$\langle 2, [.42, .42] \rangle$	0.65	30000	.51
Mr. Ganguly	28	$\langle 28, .35 \rangle$	$\langle 28, [.35, .35] \rangle$	0.59	4	$\langle 4, .48 \rangle$	$\langle 4, [.48, .48] \rangle$	0.69	30500	.59
Mr. Roy	31	$\langle 31, .44 \rangle$	$\langle 31, [.44, .44] \rangle$	0.66	6	$\langle 6, .55 \rangle$	$\langle 6, [.55, .55] \rangle$	0.74	32000	.66
Mr. David	48	$\langle 48, .94 \rangle$	$\langle 48, [.94, .94] \rangle$	0.97	16	$\langle 16, .87 \rangle$	$\langle 16, [.87, .87] \rangle$	0.93	56000	.93
Mr. Maity	54	$\langle 54, .88 \rangle$	$\langle 54, [.88, .88] \rangle$	0.94	20	$\langle 20, 1 \rangle$	$\langle 20, [1, 1] \rangle$	1	75900	.94
Mr. Das	46	$\langle 46, .88 \rangle$	$\langle 46, [.88, .88] \rangle$	0.94	15	$\langle 15, .84 \rangle$	$\langle 15, [.84, .84] \rangle$	0.92	55400	.92
Mr. Ahuja	47	$\langle 47, .91 \rangle$	$\langle 47, [.91, .91] \rangle$	0.95	14	$\langle 14, .81 \rangle$	$\langle 14, [.81, .81] \rangle$	0.9	54800	.9
Mr. Sharma	38	$\langle 38, .65 \rangle$	$\langle 38, [.65, .65] \rangle$	0.81	11	$\langle 11, .71 \rangle$	$\langle 11, [.71, .71] \rangle$	0.84	44500	.81
Mr. Kundu	36	$\langle 36, .59 \rangle$	$\langle 36, [.59, .59] \rangle$	0.77	10	$\langle 10, .68 \rangle$	$\langle 10, [.68, .68] \rangle$	0.82	40000	.77
Mr. Dutta	50	$\langle 50, 1 \rangle$	$\langle 50, [1, 1] \rangle$	1	18	$\langle 18, .94 \rangle$	$\langle 18, [.94, .94] \rangle$	0.97	78900	.97
Mr. Mondal	49	$\langle 49, .97 \rangle$	$\langle 49, [.97, .97] \rangle$	0.98	17	$\langle 17, .90 \rangle$	$\langle 17, [.90, .90] \rangle$	0.95	78200	.95
Mr. Bose	53	$\langle 53, .91 \rangle$	$\langle 53, [.91, .91] \rangle$	0.95	19	$\langle 19, .97 \rangle$	$\langle 19, [.97, .97] \rangle$	0.98	79100	.95
Mr. Gayen	51	$\langle 51, .97 \rangle$	$\langle 51, [.97, .97] \rangle$	0.98	21	$\langle 21, .97 \rangle$	$\langle 21, [.97, .97] \rangle$	0.98	80200	.98
Mr. Guha	52	$\langle 52, .94 \rangle$	$\langle 52, [.94, .94] \rangle$	0.97	25	$\langle 25, .84 \rangle$	$\langle 25, [.84, .84] \rangle$	0.92	80000	.92
Mr. Ghosh	59	$\langle 59, .74 \rangle$	$\langle 59, [.74, .74] \rangle$	0.86	33	$\langle 33, .58 \rangle$	$\langle 33, [.58, .58] \rangle$	0.76	120000	.76

We may note that since **uncertain query 2** has two fuzzy attributes *Age* and *Experience*, so similarity measure of tuples (μ) is the intersection of the corresponding similarity measures of fuzzy attributes *Age* and *Experience*.

As the query is tested for different **threshold** or α -cut values given by the decision maker, the following results are generated:

Table 8: Resultant tuples of *EMP* relation after processing uncertain query 2 using fuzzy set at threshold value $\alpha = 0.95$

Name	Age	Exp	Sal
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Bose	53	19	79100
Mr. Gayen	51	21	80200

Table 9: Resultant tuples of *EMP* relation after processing uncertain query 2 using fuzzy set at threshold value $\alpha = 0.9$

Name	Age	Exp	Sal
Mr. David	48	16	56000
Mr. Maity	54	20	75900
Mr. Das	46	15	55400
Mr. Ahuja	47	14	54800
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Bose	53	19	79100
Mr. Gayen	51	21	80200
Mr. Guha	52	25	80000

Solution using Vague Set:

Similarly, again using **Algorithm 1** and **Definition 2.3.1** respectively, the vague representation of *EMP* relation for **uncertain query 2** may be obtained as follows:

Table 10: Similarity Measure of tuples of *EMP* relation with respect to uncertain query 2 using vague set

Name	Age	Vague rep. of fuzzy data of fuzzy attribute Age	μ_1 S.M. of Age with $\langle 50, [1,1] \rangle$	Exp	Vague rep. of fuzzy data of fuzzy attribute Exp	μ_2 S.M. of Exp with $\langle 20, [1,1] \rangle$	Sal	μ
Mr. Smith	25	$\langle 25, [.26,.33] \rangle$	0.51	2	$\langle 2, [.42,.46] \rangle$	0.65	30000	.51
Mr. Ganguly	28	$\langle 28, [.35,.39] \rangle$	0.59	4	$\langle 4, [.48,.50] \rangle$	0.69	30500	.59
Mr. Roy	31	$\langle 31, [.44,.49] \rangle$	0.66	6	$\langle 6, [.55,.59] \rangle$	0.74	32000	.66
Mr. David	48	$\langle 48, [.94,.97] \rangle$	0.96	16	$\langle 16, [.87,.90] \rangle$	0.93	56000	.93
Mr. Maity	54	$\langle 54, [.88,.95] \rangle$	0.92	20	$\langle 20, [1,1] \rangle$	1	75900	.92
Mr. Das	46	$\langle 46, [.88,.94] \rangle$	0.92	15	$\langle 15, [.84,.88] \rangle$	0.91	55400	.91
Mr. Ahuja	47	$\langle 47, [.91,.96] \rangle$	0.94	14	$\langle 14, [.81,.86] \rangle$	0.89	54800	.89
Mr. Sharma	38	$\langle 38, [.65,.67] \rangle$	0.80	11	$\langle 11, [.71,.72] \rangle$	0.84	44500	.80
Mr. Kundu	36	$\langle 36, [.59,.60] \rangle$	0.77	10	$\langle 10, [.68,.7] \rangle$	0.82	40000	.77
Mr. Dutta	50	$\langle 50, [1,1] \rangle$	1	18	$\langle 18, [.94,.98] \rangle$	0.96	78900	.96
Mr. Mondal	49	$\langle 49, [.97,.98] \rangle$	0.98	17	$\langle 17, [.90,.93] \rangle$	0.94	78200	.94
Mr. Bose	53	$\langle 53, [.91,.97] \rangle$	0.94	19	$\langle 19, [.97,.99] \rangle$	0.98	79100	.94
Mr. Gayen	51	$\langle 51, [.97,.99] \rangle$	0.98	21	$\langle 21, [.97,.98] \rangle$	0.98	80200	.98
Mr. Guha	52	$\langle 52, [.94,.98] \rangle$	0.96	25	$\langle 25, [.84,.85] \rangle$	0.91	80000	.91
Mr. Ghosh	59	$\langle 59, [.74,.78] \rangle$	0.85	33	$\langle 33, [.58,.59] \rangle$	0.76	120000	.76

The result is now tested for vague set with the same **threshold** or α -cut values. The following resultant relations are now retrieved as we process the **uncertain query 2** :

Table 11: Resultant tuples of **EMP** relation after processing **uncertain query 2** using **vague Set** at threshold value $\alpha = 0.95$

Name	Age	Exp	Sal
Mr. Dutta	50	18	78900
Mr. Gayen	51	21	80200

Table 12: Resultant tuples of **EMP** relation after processing **uncertain query 2** using **vague Set** at threshold value $\alpha = 0.9$

Name	Age	Exp	Sal
Mr. David	48	16	563000
Mr. Maity	54	20	75900
Mr. Das	46	15	55400
Mr. Dutta	50	18	78900
Mr. Mondal	49	17	78200
Mr. Bose	53	19	79100
Mr. Gayen	51	21	80200
Mr. Guha	52	25	80000

It may be observed from **Tables 8** and **11** from that the resultant sets of the **uncertain query 2** for both fuzzy and vague data are different for the threshold value $\alpha = 0.95$ and vague set generates finer result. Here the vague **SQL** has not retrieved the tuples of Mr. Mondal and Mr. Bose whose age and experience both are not so close to the query. Similarly, when the same query is tested with α -cut value **0.9**, **Tables 9** and **12** again confirm that the vague **SQL** certainly gives better result than fuzzy **SQL**. The vague **SQL** has not retrieved the tuple of Mr. Ahuja with age 47 and experience 14.

5. CONCLUSIONS:

In this paper, we have presented an algorithm that generates the fuzzy or vague representation of the attributes with respect to a given uncertain query. The proposed algorithm yields an attribute independent membership function for calculation of the membership values and is free from several anomalies found in other membership functions discussed in literature [13-15]. The time complexity of the said algorithm is also presented. A comparative analysis for processing imprecise queries using fuzzy sets and vague sets has been performed using an Employee database. The present study confirms that a vague relational database model may be more fruitful in processing uncertain queries than the corresponding fuzzy model.

REFERENCES:

1. L. A. Zadeh, "Fuzzy Sets", Information and Control, Vol. 8, No. 3, pp. 338-353, 1965.
2. R. Intan and M. Mukaidono, "Fuzzy functional dependency and its application to approximate data querying", in Proceedings of international Database Engineering and Applications Symposium, pp. 47-54, 2000.
3. Y. Takahashi, "Fuzzy database query languages and their relational completeness theorem", IEEE Transactions on Knowledge and Data Engineering, Vol. 5, pp. 122-125, 1993.
4. P. Bosc and O. Pivert, "SQLF: a relational database language for fuzzy querying", IEEE Transaction on Fuzzy Systems, Vol. 3, No. 1, pp. 1-17, 1995.
5. H. Nakajima et al., "Fuzzy Database Language and Library- Fuzzy Extension to SQL", Second IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 477-482, 1993.
6. Moreau A., Pivert O. and Smits G., "Fuzzy Query by example.", 33rd ACM / SIGAPP Symposium On Applied Computing (SAC-2018), France, 2018.
7. Srivastava A., et al., "Fuzzy Query: An Impression in Query processing", Proceedings of IEEE Sponsored International Conference on Advancement in Computer Engineering and Information Technology (ACEIT 2016), 2016.
8. W. L. Gau and D. J. Buehrer, "Vague Sets", IEEE Trans. Syst. Man, Cybernetics, Vol. 23, No. 2, pp.610-614, 1993.

9. A. Lu and W. Ng, "Vague Sets or Intuitionist Fuzzy Sets for Handling Vague data: Which One Is Better?", *Lecture Notes in Computer Science*, Vol. 3716, pp. 401-416, 2005.
10. F. Zhao and Z. M. Ma, "Vague Query Based on Vague Relational Model", *AISC 61*, pp. 229-238, Springer-Verlag Berlin Heidelberg, 2009.
11. A. K. Dutta, S. Idwan and R. Biswas, "A Study of Vague Search to Answer Imprecise Query", *International Journal of Computational Cognition*, Vol. 7, No. 4, pp. 63-69, 2009.
12. J. Mishra, "An Extension of Fuzzy Relational Database Model into Vague Relational Database Model", *Ph.D. Thesis*, West Bengal University of Technology, 2014.
13. J. Mishra and S. Ghosh, "Uncertain Query Processing using Vague Set or Fuzzy Set : Which One Is Better?", *International Journal of Computers Communications & Control*, Vol. 9, No. 6, pp. 730-740, 2014.
14. Raju and A. K. Majumdar, "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database system", *ACM Transactions on Database Systems*, Vol. 13, No. 2, pp. 129-166, 1988.
15. Z. M. Ma and X. Meng, "A Knowledge-Based Approach for Answering Fuzzy Queries over Relational Databases", *LNAI 5178*, pp. 623-630, Springer-Verlag Berlin Heidelberg, 2008.
16. S. M. Chen, "Similarity Measure between Vague Sets and between Elements", *IEEE Trans. Systems. Man and Cybernetics*, Vol. 27, No. 1, pp. 153-158, 1997.
17. D. H. Hong and C. Kim, "A Note on Similarity Measures between Vague Sets and between Elements", *Information Sciences*, Vol. 115, pp. 83-96, 1999.
18. F. Li and Z. Xu, "Measures of Similarity between Vague Sets", *Journal of Software*, Vol. 12, No. 6, pp. 922-927, 2001.
19. A. Lu and W. Ng, "Managing Merged Data by Vague Functional Dependencies" Springer-Verlag Berlin Heidelberg LNCS 3288, pp. 259-272, 2004.